

<https://helda.helsinki.fi>

YSO: Avoin ja monikielinen perusta yhteentoimivalle sisällönkuvailulle

Niininen, Satu Maaret

2017-12-29

Niininen , S M , Nykyri , S K & Suominen , O I 2017 , ' YSO: Avoin ja monikielinen perusta yhteentoimivalle sisällönkuvailulle ' , Informaatiotutkimus , Vuosikerta. 36 , Nro 3-4 , Sivut 42-50 . <https://doi.org/10.23978/inf.68904>

<http://hdl.handle.net/10138/232476>

<https://doi.org/10.23978/inf.68904>

cc_by_nc_sa

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

KATSAUS

YSO: Avoin ja monikielinen perusta yhteentoimivalle sisällönkuvailulle

Satu Niininen

Kansalliskirjasto

satu.niininen@helsinki.fi

<https://orcid.org/0000-0001-6528-4743>

Susanna Nykyri

Helsingin yliopiston kirjasto

susanna.nykyri@helsinki.fi

<https://orcid.org/0000-0002-5018-5176>

Osma Suominen

Kansalliskirjasto

osma.suominen@helsinki.fi

<https://orcid.org/0000-0003-0042-0745>

Linkittyvässä ja globaalissa ympäristössä tarvitaan ratkaisuja, jotka tuovat yhteen eri paikoissa, eri aikoina ja eri kielillä kuvailtuja aineistoja yli kieli- ja kulttuurirajojen. Avoin tiede -liikkeessä on kiinnitetty paljon huomiota julkaisujen ja yhä enemmän myös tutkimusdatan avoimuuteen, mutta metatietojen avoimuuteen ja yhteentoimivuuteen kiinnitetään harvemmin huomiota. Kuitenkin metatieto voi olla avointa silloinkin, kun kuvailtava aineisto itsessään ei voi sitä olla esimerkiksi tutkimusaineiston sensitiivisyydestä tai julkaisun kustantajan rajoituksista johtuen. Tässä kirjoituksessa keskitymme monikielisyyden haasteisiin kuvailevan metatiedon välineiden laadinnassa. Ajankohtaiskatsaus pohjautuu osittain lyhennettynä ja päivitettynä aiemmin julkaistuun artikkeliin (Niininen, Nykyri & Suominen 2017), mutta tässä tekstissä kuvaamme lisäksi Kansalliskirjastossa tehtävän kehitystyön uusia ulottuvuuksia.

Asiasanat: sisällönkuvailu; monikielisuus; yhdistetty avoin tieto; ontologiat; yhteentoimivuus; tiedonhallinta

Artikkeli on lisensoitu Creative Commons Nimeä-EiKaupallinen-JaaSamoin 4.0 Kansainvälinen -lisenssillä

Pysyvä osoite: <https://doi.org/10.23978/inf.68904>

Johdanto

Kansainvälisesti tarkasteltuna tieteellisten julkaisujen määrän on arvioitu kaksinkertaistuvan noin 15 vuoden aikajaksolla (Bornmann & Mutz, 2015; Larsen & Ins, 2010; Price, 1967). Suomalaisten yliopistojen opettajien ja tutkijoiden julkaisutuottavuus ei ole kuitenkaan ratkaisevasti kasvanut sen jälkeen, kun julkaisut otettiin rahoitusmallin kriteeriksi vuonna 2007 (Pölönen & Auranen, 2017). Tämän lisäksi on kuitenkin syytä huomata myös muun tutkimuksessa käytettävän aineiston, kuten tutkimusdatan, kuvailun tarve. Myös media tuottaa paljon kuvailua. Kuvailtavan aineiston alati kasvava laajuus kasvattaa myös hyvin tuotetun ja linkittyvän sekä avoimen kuvailutiedon merkitystä, mikä parantaa kuvailtavien kohteiden löydettävyyttä. On myös syytä muistaa, että avoimeen tieteeseen sisältyvät paitsi avoin lopputuotos, niin myös sen laadintaa ohjaavien periaatteiden ja ohjeistusten avoimuus.

Avointen kuvailuvälineiden ja toimintatapojen sekä ohjeistusten rooli korostuu entisestään, kun metatietoa ovat yhä useammin tuottamassa myös muut kuin sisällönkuvailun ammattilaiset eli mm. toimittajat, tutkijat, lukijat ja tiedonhakijat. Avoimen tieteen ja tutkimuksen käsikirjassa (ATT-hanke, 2017) muistutetaan tutkijoita seuraavista asioista:

- Huolehdi kaikkien aineistojesi ja julkaisujesi mahdollisimman laajasta kuvailusta ja metatietojen avoimuudesta.
- Aikaan ja paikkaan liittyvät tiedot ovat myös tärkeitä, ilmoita ne ohjeiden mukaan. Aineistojen säilytys- tai hakupalvelut usein ohjaavat metatietojen täyttämisessä.
- Käytä oman alasi terminologiaa ja asiasanoja.
- Muista myös vapaamuotoinen ja ainakin osittain yleistajuinen kuvaus: mistä on kysymys? (description).
- Suosi valmiita sanastoja, ontologioita ja yleisessä käytössä olevia termejä löytävyyden takaamiseksi.
- Lisenssi/käyttöehdot ovat tärkeä metatieto, koneluettava formaatti kuten CC-lisenssi on suositeltava. Käytä mahdollisimman sallivaa lisenssiä. (emt.)

Käyttämällä aiheidensa kuvailussa Finto-palvelun ontologioita, tutkija huolehtii samalla osaltaan paitsi yhteentoimivien, niin myös avointen kuvailevien metatietojen käytöstä. Tällaista tarkoitusta palvelee Kansalliskirjaston Yleinen suomalainen ontologia YSO (ks. Kansalliskirjasto, 2017b). YSON kehittämistyö pyrkii olemaan mahdollisimman läpinäkyvää, ja sen sisältöön voivat vaikuttaa kaikki toimi-

jat mm. YSON ehdotusjärjestelmän kautta (ks. Kansalliskirjasto, 2017a). Metatietojen tuottaminen voidaankin nykyään nähdä jopa yhtenä välttämättömänä tutkijataitona.

Suomessa tiedejulkaiseminen on vahvasti kansainvälistä ja myös sen metatietojen käytäntöjen tulee olla kansainvälisiä. Yksi osa tätä on monikielisyys. Esimerkiksi Helsingin yliopistossa (Helsingin yliopisto, 2017) ruotsinkielisten julkaisujen osuus on pysynyt 2,4 prosentin tasolla, mutta suomenkielisten julkaisujen osuus on hieman pienentynyt ollen nyt hieman vajaa kolmannes julkaisuista (v. 2013 31 % vs. 2016 29 %). Vuonna 2016 kaikista vertaisarvioituista julkaisuista kansainvälisiä oli 87 prosenttia (6088). Tutkimuksen kansainvälistyminen näkyy englanninkielisten julkaisujen osuuden tasaisena vuosittaisena kasvuna: vuonna 2013 näiden osuus oli noin 63 prosenttia ja vuonna 2016 66 prosenttia. Yhtenä syynä englanninkielisen julkaisemisen kasvuun on kansainvälisen opetus- ja tutkimushenkilöstön osuuden kasvu. (Emt.)

Kansalliset ja kansainväliset näkökulmat ovat metatietojen tuottamisessa usein keskenään yhteneväiset. Silloinkin kun alkuperäinen julkaisu on tuotettu kansallisella kielellä, sitä haetaan, siitä keskustellaan ja siihen viitataan muunkin kielisissä yhteyksissä, ja sen metatietojen tulee linkittyä myös kansainvälisiin tietovarantoihin. Tällöin monikielisten ja kattavien metatietojen merkitys korostuu. Mutta vaikka monikielisuuden ja avoimuuden tarpeesta vallitsikin yksimielisyys, ei sen toteuttaminen ole vailla haasteita.

Finto-palvelu ja Yleinen suomalainen ontologia YSO

Finto on Kansalliskirjaston ylläpitämä suomalainen sanasto- ja ontologiapalvelu, joka mahdollistaa sanastojen julkaisun ja selailun. Palvelu tarjoaa myös rajapinnat sanastojen ja ontologioiden hyödyntämiseen muissa ohjelmistoissa. Kehitystyöstä vastaa opetus- ja kulttuuriministeriön ja valtiovarainministeriön rahoituksella Kansalliskirjastossa toimiva Finto-projekti. Palvelun kehityksessä on hyödynnetty Aalto-yliopiston ja Helsingin yliopiston Semanttisen laskennan ryhmän FinnONTO-projektin vuosina 2003–2012 tekemää pohjatyötä (Hyvönen, 2014; Lappalainen, 2014).

Finto-palvelun keskeisimpänä sisältönä toimii Yleinen suomalainen ontologia YSO. YSO on rakennettu suomalaisen kulttuuripiirin sisällönkuvailutarpeiden ja käsitteistön pohjalta, ja se pohjautuu yleiseen suomalaiseen asiasanastoon (YSA) sekä yleiseen ruotsinkieliseen tesaurukseen (Allärs). YSO on koostettu muokkaamalla YSAn ja Allärsin termistöstä yhtenäinen kokonaishierarkia, joka sisältää käsitteiden välisiä suhteita koneluettavassa muodossa. Hierarkisoinnin yhteydessä on tehty siirtymä termitasolta käsitetasolle purkamalla käsitteiden monimerkityksisyyttä

ja ankkuroimalla ne aiempaa täsmällisempiin merkityksiin. Samalla käsitteille on li-
sätty myös englanninkieliset vastineet ja linkityksiä muihin sanastoihin, toistaiseksi
pääasiassa Yhdysvaltain Kongressin kirjaston LCSH-sanastoon.

Standardit

YSOssa käytetty käsitte pohjainen mallinnustapa noudattaa kansainvälistä ISO-
standardia (ISO 25964-1:2011, 2011), joka määrittelee sekä yksi- että monikielisten
tesaurusten laadinnassa käytettävät periaatteet. Aiemmista termeihin vahvasti no-
jaavista tesaarusstandardeista poiketen lähestymistapa on käsitte pohjainen, eli kes-
kiössä ovat ajatusyksiköt eli käsitteet, jotka yksilöidään sekä termein että yksilöivien
URI-tunnistein. Käsitteiden tunnisteet pyritään myös säilyttämään pysyvinä, vaika
ka termit ja sanaston rakenne ajan kanssa muuttuisivat. Hierarkkiset ja assosiatii-
viset suhteet luodaan käsitteiden, ei termien, välille, joten monikielisessä sanastos-
sa rakenne on yhteinen kaikille kielille. YSON julkaisuversion RDF-pohjainen tie-
tomalli perustuu W3C:n määrittelemään Simple Knowledge Organization System
(SKOS)-malliin (Isaac & Summers, 2009), joka on yleiskäyttöinen, melko yksinker-
tainen esitystapa tesauruksille, luokituksille ja muille sanastoille, jotka voidaan esit-
tää joukkona käsitteitä ja termejä. SKOS-malli määrittelee myös joukon linkitys-
suhteita, joiden avulla käsitteet voidaan linkittää muiden sanastojen käsitteisiin.

YSON käsitteenmuodostus perustuu suomen ja ruotsin kielisiin, jotka ovat kes-
kenään tasavertaisessa asemassa käsitteitä muodostettaessa. Englannin kielellä on
käännöskielen asema, joten se ei vaikuta käsitteenmuodostukseen.

Monikielisyyden haasteet

Monikielisen sanaston rakentamisessa on omat haasteensa. Käsitteiden kantamat
merkitykset ovat eri kielissä vain harvoin täysin symmetrisiä, sillä jokainen kult-
tuuri jäsentää maailmaa eri näkökulmasta. Koska YSO on rakentunut suomalaisen
sisällönkuvailun tarpeista, se sisältää suomessa käytettävää käsitteistöä, jolle ei löy-
dy suoria vastineita englanniksi eikä aina myöskään (suomen)ruotsiksi. Tällaisia
ovat esimerkiksi Suomen historian tapahtumat ja ilmiöt (*mäkitupalaiset*, *isojako*,
Ahvenanmaan kysymys) sekä yhteiskuntaan ja kulttuuriin liittyvä käsitteistö (*tal-
koot*, *neuvolat*, *erillistalot*, *sosiaalikasvattajat*). Lisäksi oman erityisryhmänsä muo-
dostavat vanhentuneet termit tai käsitteet, jotka ovat joko muuttuneet toisen ni-
misiksi (*lääkintävoimistelu*, nyk. *fysioterapia*) tai joita käytetään enää viittaamaan
oman aikansa ilmiöihin (*suostuntavero*). (ks. Kansalliskirjasto, 2017b)

Koska ontologian rakentamisen lähtökohtana on ollut ajatus, että jokainen kä-
site on oma ajatusyksikkönsä, myös niistä käytettävät nimitykset ovat yksilöllisiä.

Kun eri kielten käsitteistöt eroavat toisistaan, joudutaan tekemään kompromisseja yhtenäisen hierarkian rakentamisessa. Tämä näkyy esimerkiksi käsitteessä *tädit*, jolla on ruotsinkielisenä käytettävänä terminä fraasi *fastrar och mostrar*, sillä käsitteellä on ruotsiksi kaksi erillistä termivastinetta (ks. taulukon 1 esimerkki 1.1). Vastaava ratkaisu toisin päin on tehty ruotsin käsitteessä *skyltar*, jonka suomenkielisenä vastineena on fraasi *kilvet ja kyltit* (taulukon 1 esimerkki 1.2). Myös käsitteellä *hyvinvointi* on ruotsiksi kaksi toisistaan poikkeavaa merkitystä *välmaende* ja *välfärd*, ja tässä tapauksessa on päädytty jakamaan merkitykset kahdeksi erilliseksi käsitteeksi, joiden suomenkieliset terminuodot on erotettu toisistaan sulkutarkenteiden avulla (taulukon 1 esimerkki 1.3). Käsitteiden jakamiseen ja yhdistämiseen liittyvät ratkaisut vaikuttavat luonnollisesti myös siihen, mihin ne sijoittuvat ontologian hierarkiassa. Koska suomi ja ruotsi ovat molemmat tasa-arvoisia käsitteenmuodostuskieliä, pyritään käsittejakoa tekemään joustavasti ja tasapuolisesti molempien kielten ehdoilla. Mikäli pelkästään englanninkielinen käännös on monitulkintainen tai muuten ongelmallinen, ei käsitteen rajaukseen tai hierarkisointiin tehdä muutoksia vaan etsitään soveltuva käännösratkaisu muuta kautta, esimerkiksi juuri sulkutarkenteiden tai erillisten käyttöhuomautusten avulla (taulukon 1 esimerkki 1.4).

Taulukko 1: Monikielisyyden haasteet

	Käsite suomeksi	Käsite ruotsiksi	Englanninkielinen käännös
1.1	tädit	fastrar och mostrar	aunts
1.2	kilvet ja kyltit	skyltar	signboards and plates
1.3	hyvinvointi (terveydellinen)	välmaende	well-being
1.3	hyvinvointi (taloudellinen)	välfärd	prosperity
1.4	tasa-arvo	jämställdhet	equality (values)
1.4	yhdenvertaisuus	jämlikhet	equality (fundamental rights)
1.5	imeytyminen	uppsugning	absorption
1.5	absorptio	absorption	absorption

Kaikille käsitteille ei välttämättä ole löydettävissä täysin yksiselitteistä ja tyhjentävää vastinetta muilla kielillä, jolloin olemassa olevista vaihtoehtoista joudutaan tekemään valinta sen perusteella, mikä vaikuttaa käyttökontekstin kannalta parhaiten soveltuvalta. Jos käsitteelle löytyy useita tasavertaisia käännösvastineita, voidaan täsmällisin tai toissijaisesti yleisin näistä valita käytettäväksi termiksi ja muut ohjaustermeiksi. Tällöin korvaut termit auttavat käsitteen merkityksen rajaamisessa eivätkä ne välttämättä ole toistensa tai suositeltavan termin synonyymeja. Lisäksi englanninkielisissä vastineissa on jouduttu tekemään joitakin poikkeuksia yksilöivien termien käyttämiseen tapauksissa, joissa käsitteiden välille ei englanniksi tehdä eroa, esimerkiksi käsitteet *imeytyminen* ja *absorptio* kääntyvät molemmat *absorption* (taulukon 1 esimerkki 1.5).

Kääntämisestä linkitykseen

Englanninkielisen käännöksen kautta YSOa on linkitetty myös kansainvälisiin sanastoihin, joista ensimmäisenä työn kohteeksi on otettu Library of Congress Subject Headings (LCSH). Käytännössä tämä on toteutettu etsimällä LCSH:sta ja YSOsta toisiaan vastaavia käsitteitä, ja luomalla näiden välille linkitys. Haasteita linkitykseen tuo se tosiasia, että sanastot eivät pelkästään sisällä erilaista käsitteistöä vaan eroavat myös rakenteensa ja käytäntöjensä osalta. Esimerkiksi toisin kuin YSO, LCSH sisältää runsaasti asiasanojen ketjuttamista tukevia rakenteita.

Linkityksen tarkoituksena on tarjota tiedonhakijalle YSO-käsitteiden tiedonhaullisia vastineita LCSH:sta. Tällöin lähtökohtana pidetty sitä, että linkityksen voi tehdä silloin, kun vastaavuus on niin vahva, että käsitteitä voisi käyttää toistensa vastineina sisällönkuvailussa. Tällöin voidaan olettaa, että tietokannasta tehty asiasanahaku tarttuisi kumman tahansa sanaston käsitteellä pääpiirteittäin samoihin aineistoihin ja tiedonhakija saisi molemmilla kielillä paljolti toisiaan vastaavat hakutulokset (taulukon 2 esimerkki 2.1).

Suurin osa LCSH-linkityksistä tehdään skos:closeMatch-suhteella. Se on tarkoitettu yhdistämään kaksi käsitettä, jotka ovat riittävän samankaltaiset, että niitä voidaan käyttää toistensa asemesta tiedonhakuovelluksissa (Isaac & Summers, 2009).

CloseMatch-suhde ei siis tarkoita, että käsitteiden välillä olisi täydellinen vastaavuus kaikissa käyttötilanteissa. Tämä tuo käsitteiden väliseen ekvivalenssiin hiekan liikkumavaraa, sillä linkitys voidaan tehdä myös käsitteille, joiden katsotaan olevan riittävän lähellä toisiaan. Tämän lisäksi käytössä ovat skos:broadMatch- ja skos:narrowMatch-suhteet, joita käytetään täydentävästi silloin, kun vastaavuus on selvästi hierarkkinen, eli yhden sanaston käsite ilmaistaan toisessa sanastossa usean käsitteen yhdistelmällä. Nämä suhteet on otettu käyttöön vasta linkitysprojektin loppuvaiheessa, joten niiden käyttö ei vielä toistaiseksi ole täysin yhtenäistä läpi koko ontologian.

Linkitysten keskeisimpänä haasteena on riittävän vastaavuuden määrittäminen: milloin käsitteet ovat merkityksiltään niin läheisiä, että niiden voidaan katsoa vastaavan toisiaan? Erityisen ongelmallisia ovat esimerkiksi käsitteet, jotka ovat merkityssisällöltään vain osittain päällekkäisiä tai joissa yhden käsitteen merkityssisältö ilmaistaan toisessa sanastossa useamman käsitteen osittaisella yhdistelmällä (taulukon 2 esimerkki 2.2). Yhtä pulmallisia ovat myös tapaukset, joissa lähde- ja kohdesanastojen käsitteistö liikkuu eri tarkkuustasolla (taulukon 2 esimerkki 2.3), eikä yhteistä yleistermiä löydy. Myöskään termitason vastaavuus ei välttämättä ole tae käsitetason vastaavuudesta, sillä sama termi on voitu ottaa eri sanastoihin täysin eri merkityksissä (taulukon 2 esimerkki 2.4).

Linkitys jätetään kokonaan tekemättä silloin kun soveltuvaa ei löydy lainkaan

tai jos lähin potentiaallinen vastine on niin kaukainen, ettei voi ajatella sen tuovan tiedonhakuun lisäarvoa (taulukon 2 esimerkki 2.5). Ilman linkitystä jäävät käsitteet eivät suinkaan edusta yksinomaan suomalaisen kulttuuripiirin käsitteistöä vaan joukossa on myös varsin globaaleja ilmiöitä, kuten *synergia* ja *paronit*.

Taulukko 2: Kääntämisestä linkitykseen

	Käsite suomeksi	Englanninkielinen käännös	Lähin vastaava käsite LCSH:ssa	Linkityssuhde
2.1	alkuopetus	elementary education Note: Refers to the first two years of primary school education: the 1st and 2nd grades	education, primary Note: Here are entered works on education limited to grades one through three or four.	closeMatch
2.2	avaruuslennot	space flights	space flights Note: Here are entered works giving accounts of events and experiences during specific manned space flights.	closeMatch
			space flight Note: Here are entered works on the physics and technical details of locomotion beyond the earth's atmosphere.	closeMatch
2.3	saippuaopperat	soap operas	radio soap operas television soap operas	narrowMatch
2.4	fundamentalismi	fundamentalism	Fundamentalism Note: Here are entered works on the modern Protestant movement originating in early 20th century America...	narrowMatch
2.5	helsinkiläisyys	regional identity of Helsinki residents	identity (psychology)	–

Ongelmatapausten yhteydessä on aina punnittava, onko vastaavuus niin vahva, että linkitys on perusteltua. Työn alkuvaiheessa keskeisimpänä haasteena onkin ollut määrittää, millaisella tarkkuudella vastaavuus määritellään ja millaisia linjauksia linkityksissä noudatetaan, jotta ne toteutuisivat kautta linjan johdonmukaisina. Kaikille käsitteille ei löydy linkitysvastinetta lainkaan, ja näissä tapauksissa käsite jää ilman linkitystä. Toistaiseksi noin 44 prosenttia YSON käsitteistä on saatu linkitettyä LCSH-käsitteisiin, ja luku tulee todennäköisesti vielä kasvamaan sen myötä, kun soveltuvia narrow- ja broad-linkityksiä takautuvasti täydennetään.

Kohti (puoli)automaattista sisällönkuvailua

YSON käsitteepohjainen, monikielinen ja linkitetty rakenne luo erinomaisen pohjan muistiorganisaatioiden sisällönkuvailu- ja tiedonhakuprosessien kehittämiselle ja

automaatiolle. Nykyisiä kuvailutyön prosesseja voidaan tehostaa ja uusia mahdollisuuksia avautuu, kun sanasto nähdään termijoukon sijaan kokoelmana käsitteitä, joilla on yksilöivät, julkiset URI-tunnisteet ja joihin liittyy rakenteista metatietoa ja linkkejä.

Kirjastoissa tehdään paljon kopioluettelointia, missä ulkopuolisesta lähteestä saatuja tietueita kopioidaan osaksi omaa kirjastotietokantaa. Kopioinnin yhteydessä tietueet on kuitenkin sovittava paikallisiin käytäntöihin. Esimerkiksi tietueessa esiintyvät, kuvailusääntöjen mukaiset vakiofraasit (esim. toimijan roolit kuten “kirjoittaja” ja “kääntäjä” ja teostyytit kuten “teksti”) on muutettava suomenkieliseen kirjoitusasuun, mikä yleensä onnistuu helpoimmin muunnostaulukoiden avulla.

Sisällönkuvailun osalta tilanne ei kuitenkaan ole niin yksinkertainen. Kopioitavissa tietueissa on harvoin käytetty suomalaisia sanastoja, mutta LCSH:lla tehdyt sisällönkuvailut ovat varsin yleisiä. YSON LCSH-linkit tarjoavat mahdollisuuden muuntaa osa näistä LCSH-aiheista YSO-aiheiksi. Haasteita tulee kuitenkin siitä, että sanastojen kokoeron ja vain osittaisen päällekkäisyyden vuoksi YSON LCSH-linkit kattavat vain pienen osan (noin 3,4 %) kaikista LCSH:n yli 400 000 käsitteestä. Lisäksi LCSH-aiheet ilmaistaan tyypillisesti asiasanaketjuina, kun taas YSOa käytettäessä ketjutusta pyritään välttämään. Alustavissa kokeiluissa Melinda- ja Helka-tietokantojen tietueilla on kuitenkin saatu lupaavia tuloksia: noin kahdelle kolmannekselle niistä tietueista, joilla oli LCSH-aiheita, pystyttiin YSON LCSH-linkkien kautta tuottamaan vähintään yksi YSO-aihe. Kokeiluja on tarkoitus jatkaa ja verrata myös näin tuotettujen YSO-sisällönkuvailujen laatua alkuperäisiin LCSH-sisällönkuvailuihin.

Toinen kehittämisen kohde on sisältöpohjainen automaattinen sisällönkuvailu. Työkaluja ja kaupallisia tuotteita automaattiseen sisällönkuvailuun on varsin runsaasti tarjolla, mutta vain harvat niistä soveltuvat suomenkielisille dokumenteille. Käsitepohjainen, tunnisteilla varustettu sanasto toimii kiintopisteenä, jonka ympärille automaattista sisällönkuvailua on hyvä kehittää. Eri työkaluja voidaan myös verrata keskenään, kun lähtökohdaksi asetetaan se, että kaikkien työkalujen on kyettävä tuottamaan YSO-käsitteistä koostuva sisällönkuvailu.

Kansalliskirjastossa on vireillä automaattiseen sisällönkuvailuun soveltuvien työkalujen testaus, jota tehdään erityyppisillä suomenkielisillä dokumenteilla. Kansalliskirjastossa kehitetään myös omaa, tilastolliseen malliin pohjautuvaa Annifytyökalua (ks. Annif.org, 2017). Työkalun ensimmäinen prototyyppi antoi lupaavia tuloksia, joten sen kehitystä on päätetty jatkaa.

Lopuksi

Sisällönkuvailun tulee ottaa huomioon erilaiset kuvailijat ja tiedonhakijat paitsi yli kieli- ja kulttuurirajojen, niin myös yli eri aikakausien. Pitkäaikaissäilytyksen arvo on pitkäaikaissäilytävyydessä, ja se vaatii sisällönkuvailussa käytettävien menetelmien ja terminologian jatkuvaa päivitystyötä ja mm. pysyvien tunnisteiden käyttöä. Tulevaisuudessa kirjastojen tuottama metatieto ja sitä hyödyntävät välineet ovat yhä useammin luonteva osa Internetissä toimimista, ilman että käyttäjien tarvitsee erikseen opetella niiden käyttöä. Avointen ja ajantasaisten työkalujen avulla pystytään paremmin vastaamaan myös muuttuvan informaatioympäristön tarpeisiin ja palvelemaan niin erilaisia käyttöjä kuin käyttäjiäkin.

Viitteet

- Annif.org. (2017). Annif - automated subject indexing using Finna as a corpus. <http://annif.org/> (luettu 17.12.2017)
- ATT-hanke. (2017). Avointen tieteen käsikirja tutkijoille ja tutkimusorganisaatioille. <https://avointiede.fi/www-kasikirja> (luettu 17.12.2017)
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222. <https://oadoi.org/10.1002/asi.23329>
- Helsingin yliopisto. (2017). *Katsaus vuoteen 2016 ja strategiakauteen 2013–2016*. <http://bit.ly/strategiakausi-2013-2016> (luettu 17.12.2017)
- Hyvönen, E. (2014). FinnONTO-hanke loi ontologisen perustan kansalliselle webin tietoinfrastruktuurille. *Tieteessä tapahtuu*, 32(3), 12–18. <https://journal.fi/tt/article/view/41559> (luettu 17.12.2017)
- Isaac, A., & Summers, E. (2009). SKOS Simple Knowledge Organization System Primer. *W3C Group Note*. <https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> (luettu 17.12.2017)
- ISO 25964-1:2011. (2011). ISO 25964 – the international standard for thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. <http://www.niso.org/schemas/iso25964> (luettu 17.12.2017)
- Kansalliskirjasto. (2017a). YSA:n ja YSO:n käsite-ehdotuslomake. <http://ehdotus.finto.fi/> (luettu 17.12.2017)
- Kansalliskirjasto. (2017b). YSO - Yleinen suomalainen ontologia. <https://finto.fi/yso/> (luettu 17.12.2017)
- Lappalainen, M. (2014). *Yläontologiat ja yleisen suomalaisen ontologian ylärakenne*. <http://urn.fi/URN:ISBN:978-951-51-0347-5> (luettu 17.12.2017)
- Larsen, P. O., & Ins, M. von. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575–603. <https://oadoi.org/10.1007/s11192-010-0202-z>
- Niininen, S., Nykyri, S., & Suominen, O. (2017). The future of metadata: open, linked, and multilingual – the YSO case. *Journal of Documentation*, 73(3), 451–465. <https://oadoi.org/10.1108/JD-06-2016-0084>
- Pölonen, J., & Auranen, O. (2017). Julkaisupaine suomalaisessa tiedeyhteisössä. *Informaatiotutkimus*, 36(2). <https://oadoi.org/10.23978/inf.65186>
- Price, D. J. de S. (1967). *Little science, big science... and beyond*. New York: Columbia University Press. <http://derekdesollaprice.org/little-science-big-science-full-text/> (luettu 17.12.2017)